

L'évaluation des apprentissages scolaires : une question de justesse

Micheline Bercier-Larivière

Renée Forgette-Giroux

université d'ottawa

De nombreux chercheurs déplorent l'absence d'un modèle de qualité des résultats d'évaluation des apprentissages en salle de classe. Après avoir précisé pourquoi le modèle psychométrique, avec ses critères de validité et de fidélité, ne convient pas au domaine scolaire, les auteures proposent le concept de justesse comme critère de qualité des résultats d'évaluation en salle de classe. Ce dernier tient compte de la pertinence de la tâche d'évaluation par rapport aux habiletés visées, de sa cohérence avec l'activité pédagogique, de sa transparence pour l'élève et de l'absence de désavantage circonstanciel.

Many researchers deplore the lack of a good model for learning assessment in the classroom. After clarifying why the psychometric model, with its criteria of validity and reliability, is not appropriate for the school situation, we suggest the concept of soundness as a criterion for the quality of assessment results in the classroom. Soundness takes into account the pertinence of the assessment task relative to the targeted skills, its appropriateness for the educational activity, its transparency for the student, and the absence of disadvantageous circumstances.

L'enseignement traditionnel caractérisé par la transmission de connaissances et la sélection des meilleurs élèves est graduellement disparu. Il a cédé la place à l'apprentissage dynamique visant la maîtrise d'habiletés complexes transdisciplinaires. Celles-ci sont susceptibles d'amener tous les élèves à « acquérir les connaissances et les habiletés qui assureront leur réussite dans le contexte de la mondialisation de l'économie tout en leur permettant de mener avec intégrité une vie civique et personnelle satisfaisante » (Ministère de l'Éducation et de la Formation de l'Ontario, 1997, p. 3). Parallèlement, l'évaluation des apprentissages a évolué, tentant de s'adapter aux nouvelles philosophies du système d'éducation, aux théories d'apprentissage qu'il adopte et aux missions qu'il se donne. Aussi, une partie importante de l'activité en salle de classe consiste à évaluer les apprentissages. Les résultats obtenus servent à de nombreuses décisions importantes pour l'avenir des élèves. La promotion des élèves et l'attribution des diplômes se basent également sur ces résultats. Pourtant, le personnel enseignant n'a souvent ni l'encadrement ni la formation appropriée pour assurer la qualité des résultats d'évaluation. De plus, le domaine de l'éducation ne possède pas de modèle théorique adéquat de la qualité des résultats d'évaluation des apprentissages en salle de classe, qu'elle soit formative, sommative ou

certificative; il n'a pas de schème auquel se référer de manière à arriver à des résultats justes, c'est-à-dire représentant fidèlement le développement de chaque élève. Un tel schème donnerait pourtant de la rigueur et de la cohérence à la démarche d'évaluation, faciliterait la communication entre intervenants et rendrait la formation plus simple et plus efficace. La recherche sur l'évaluation en salle de classe basée sur un modèle de qualité des résultats serait aussi plus efficace si elle s'appuyait sur une vue d'ensemble cohérente d'éléments à étudier et à analyser. Pour le moment, elle demeure souvent fragmentée et superficielle.

Dans cette optique, certains chercheurs tentent d'adapter au contexte de la salle de classe les critères de validité et de fidélité élaborés d'abord en psychologie et de les « recycler » en principes éducatifs. D'autres favorisent plutôt la définition de concepts nouveaux et de principes propres au domaine, ancrés dans le contexte de l'enseignement-apprentissage en salle de classe. D'après ces derniers, les notions de validité et de fidélité ont été galvaudées dans le contexte de l'éducation, à la faveur de l'apparence de précision, de scientificité et d'objectivité que leur conféraient les calculs statistiques. Mais au-delà de la terminologie défendue, que révèle cette argumentation au sujet des préoccupations fondamentales de qualité des résultats d'évaluation des apprentissages en salle de classe? Ne pourrait-elle pas, en fait, faire émerger les bases d'une conception de la qualité propre à l'évaluation des apprentissages scolaires?

Le présent article a pour but de mettre en évidence les éléments fondamentaux de la qualité de l'évaluation des apprentissages et leur interrelation. Il décrit sommairement les pratiques d'évaluation en salle de classe telles que les recherches empiriques les ont dépeintes jusqu'à maintenant, discute des critères de qualité élaborés en psychométrie et propose une conception de la qualité de l'évaluation en salle de classe basée sur la justesse des résultats.

LES PRATIQUES D'ÉVALUATION EN SALLE DE CLASSE

L'intérêt pour l'étude des pratiques évaluatives en salle de classe date d'une vingtaine d'années. Les recherches indiquent qu'à l'élémentaire, les enseignants consacrent en moyenne entre un tiers et un quart du temps de classe à des tests (Herman et Dorr-Bremme, 1984). Des enseignants de sciences de huitième année disent mener des activités d'évaluation plus fréquemment que toute autre activité, au détriment de démonstrations, d'expérimentations et d'activités pédagogiques diverses (Herman et Dorr-Bremme, 1984). Au secondaire, un élève ontarien de dixième année subit en moyenne une épreuve formelle à tous les deux jours; au cours d'une année scolaire, ceci équivaut à 24 heures et 18 minutes (Wilson, 1989).

Les préoccupations des enseignants sont d'abord d'accumuler des notes pour le bulletin et de confirmer les jugements déjà établis (Forgette-Giroux, Simon et Bercier-Larivière, 1996; Herman et Klein, 1996; Wilson, 1990). Souvent, évaluer veut simplement dire accorder des points aux tâches d'apprentissage assignées

et tester le contenu enseigné, c'est-à-dire vérifier ce que les élèves en ont retenu (Bercier-Larivière, 1998). Les enseignants le font sans se référer à un schème d'interprétation particulier ni à des critères précis (Macdonald, 1993; Pryor, 1992). Ils tiennent parfois compte d'éléments qui ont été ignorés lors de l'apprentissage ou qui n'ont rien à voir avec l'habileté visée. Ainsi, des élèves de troisième année ont remarqué que même si la compréhension est le critère important lors de l'apprentissage de la lecture, lors de son évaluation c'est plutôt la qualité des réponses écrites et la qualité de l'élocution qui sont retenues (Davinroy, 1994). De plus, souvent les enseignants tiennent compte du comportement en salle de classe, de l'effort ou du progrès pour décider du niveau de performance d'un élève (Bateson, 1990; Cox, 1994).

Les enseignants éprouvent des difficultés à intégrer l'évaluation à l'enseignement, font peu de rétroaction et réinvestissent rarement les résultats dans l'intervention pédagogique (Conseil supérieur de l'éducation, 1992; Stiggins et Conklin, 1992; Wilson, 1990). Ils modifient peu le cours des activités pédagogiques prévues au programme lorsque, dans une épreuve de fin d'étape, les élèves atteignent un niveau de maîtrise insatisfaisant (Bercier-Larivière, 1994). Pour les bilans, les notes proviennent habituellement de résultats d'évaluation formative combinés à ceux de fin d'étape (Bachor et Anderson, 1994; Conseil supérieur de l'éducation, 1992; Forgette-Giroux et al., 1996).

Les recherches faisant appel à des méthodologies qualitatives, plus populaires depuis le début des années 90, indiquent que les enseignants éprouvent de la difficulté à décrire leur façon d'évaluer parce qu'ils demeurent vulnérables au désir de bien paraître (Higgins et Rice, 1991; Stiggins, 1990). Conséquemment, il faut garder une certaine réserve face aux pratiques décrites à partir de toute méthode indirecte. Pourtant, grâce à l'approche directe, il a été démontré que dans des cours de sciences où l'accent est supposément mis sur les habiletés supérieures de la pensée et le développement de l'esprit critique, on évalue en fait des habiletés élémentaires comme la mémorisation et la connaissance de faits plutôt que la compréhension (Fraser et Tobin, 1991; Tobin, Butler-Kahle et Fraser, 1990). Ces recherches qualitatives permettent aussi de voir que l'évaluation des apprentissages s'apparente souvent à l'expression de l'autorité et que les notes servent parfois à assurer la discipline (Kusch, 1995).

Ces quelques exemples d'utilisation fautive de l'évaluation et des difficultés rencontrées par les enseignants ne doivent toutefois pas étonner outre mesure. Bien que l'utilisation positive de l'évaluation soit prescrite depuis de nombreuses années par les instances gouvernementales, tout comme l'utilisation de critères et de normes pour évaluer les apprentissages, peu de recherches ont vérifié la capacité des enseignants à le faire, ou confirmé le changement de leurs pratiques. Les études effectuées tant au Canada qu'aux États-Unis démontrent que jusqu'à maintenant, la formation des enseignants à l'évaluation des apprentissages n'a pas été très valorisée, ni même vue comme une compétence à développer (Biggs, 1995; Stiggins, 1991). Même si l'évaluation quotidienne des apprentissages

constitue une des tâches les plus complexes et les plus exigeantes, pour la plupart des enseignants, l'apprentissage de l'évaluation se fait par essais et erreurs, aux dépens des élèves. Au mieux, ceux qui disent avoir effleuré le sujet au cours de leurs études ne jugent pas cette formation pertinente pour les besoins quotidiens de la salle de classe (Bateson, 1990; Dempster, 1992; Rogers, 1991; Shepard, 1992; Stiggins, 1990).

Ces lacunes sur le plan de la formation des enseignants sont en grande partie attribuables à l'absence d'un cadre conceptuel de la qualité des résultats d'évaluation des apprentissages en salle de classe. En effet, bien que de nombreux manuels fassent la promotion de diverses visions de l'évaluation en salle de classe, il n'existe pas présentement de modèle conceptuel éprouvé pour structurer l'évaluation des apprentissages de manière à assurer des résultats de qualité (Biggs, 1995; Gipps, 1994; Gipps et Murphy, 1994; Mislevy, 1992; Rogers, 1991; Taylor, 1994). Cette absence, déplorée il y a déjà plus de 30 ans, n'a toujours pas été comblée. Les fondements de la mesure en psychologie — qui sert depuis longtemps, par défaut, en éducation — contiennent les raisons de son incompatibilité avec les apprentissages scolaires tels que définis actuellement.

LA PSYCHOMÉTRIE

Les réalisations de Binet-Simon et Terman pour mesurer l'intelligence humaine servent de base à la psychométrie (Trahan et Dassa, 1978). Cette conception de l'évaluation a pour prémisses que la caractéristique ou la qualité mesurée est unidimensionnelle et stable et se répartit dans la population en suivant une courbe de Gauss. Pour mesurer un élément psychologique et pouvoir le distribuer normalement dans une population donnée, la mesure doit se concentrer sur ce qui distingue les individus entre eux, dans le but de les comparer et d'en faire ressortir les différences. Cet effet de microscope sur les légères variations de l'attribut mesuré permet de le quantifier et de le répartir selon une courbe de distribution normale.

En psychologie, la capacité d'un instrument à bien quantifier un trait mesuré s'exprime par sa validité et sa fidélité. De ces deux indicateurs de qualité, la validité est l'indice de la capacité d'un outil à mesurer réellement ce qu'il prétend mesurer, selon l'utilisation que l'on veut en faire (Popham, 1990). La fidélité, pour sa part, se définit comme la constance avec laquelle une personne obtiendrait le même résultat dans des conditions semblables. Il existe des principes d'élaboration d'un instrument de mesure susceptibles de favoriser des résultats fidèles et valides (méthode *a priori*). Par exemple, des experts peuvent se prononcer sur la pertinence des items choisis pour évaluer le contenu ciblé, mais la confirmation se fait toujours à partir d'une application de l'outil en question (méthode *a posteriori*). La validité et la fidélité se mesurent par des tests d'hypothèse pour lesquels on applique un modèle mathématique précis à des

données recueillies auprès d'un groupe de sujets. Pour qu'il puisse maintenir son niveau de qualité, il est essentiel que l'outil en question soit toujours administré de manière uniforme, à des sujets en tous points identiques à la population considérée au départ (Messick, 1989).

La qualité des résultats au sens psychométrique du terme nécessite donc une variance maximale des scores et un assez grand nombre d'items indépendants les uns des autres (Lussier et Turner, 1995). Selon Taylor (1994), cette notion de qualité tient essentiellement aux différences individuelles et à notre habileté à bien les mesurer. Tout se joue sur la longueur de l'épreuve, le type d'items utilisés, la distribution des scores, les conditions d'administration standardisées et les modèles mathématiques établis pour en confirmer la validité et la fidélité.

Or l'évaluation en salle de classe a généralement recours à un petit nombre d'items, en particulier lorsqu'il s'agit d'habiletés complexes; il devient donc impossible d'obtenir une distribution variée des résultats, comme le modèle psychométrique l'exige. De plus, outre ce problème, l'enseignant n'a pas la possibilité de vérifier la validité et la fidélité des épreuves qu'il prépare en les appliquant d'abord à un autre groupe d'élèves en tous points identiques aux siens. Voilà donc deux points rendant impossible l'établissement de la qualité des résultats d'évaluation des apprentissages à la manière psychométrique. Mais quel sens devrait plutôt prendre le terme « qualité des résultats » dans une perspective éducatrice? Qu'est-ce qui détermine la qualité des résultats lorsqu'il s'agit d'évaluer des apprentissages scolaires dans le contexte de la salle de classe?

L'ÉDUMÉTRIE

L'apprentissage se définit comme « un processus d'effet plus ou moins durable par lequel des comportements nouveaux sont acquis ou des comportements déjà présents sont modifiés » (DeLandsheere, 1992, p. 20). Dans sa conception moderne, l'apprentissage diffère des traits psychologiques par sa multidimensionalité (Gipps, 1994) et sa constante évolution.

L'évaluation des apprentissages consiste à mener une démarche stimulant l'apparition d'indices observables qui en témoignent, afin d'apprécier la qualité ou l'étendue de leur développement. Les apprentissages ne peuvent qu'être estimés, toujours de manière imprécise, selon le contexte dans lequel ils sont observés (Gipps et Murphy, 1994) et appréciés selon un idéal à atteindre (Hipps, 1993). Puisque l'élève apprend constamment, les résultats d'une évaluation donnée deviennent vite périmés (Shepard, 1992).

L'intervention éducative efficace a besoin de l'évaluation pour favoriser l'aide, la communication et le renforcement (Cardinet, 1985). L'évaluation, étroitement liée au processus pédagogique, rend ce dernier plus dynamique en s'intéressant davantage au mode de pensée et au cheminement qu'aux résultats (Kulm, 1994). Sa forme doit être constructive, formative et intégrée à l'apprentissage. D'une part, elle permet de raffiner l'enseignement, par exemple d'adapter

des tâches d'apprentissage aux intérêts personnels (Harnisch et Mabry, 1993; Leder, 1992). D'autre part, elle donne régulièrement le pouls de l'apprenant, indique s'il progresse normalement et si la stratégie d'enseignement convient bien (Baron, 1991).

En 1963, Glaser a été le premier à soutenir publiquement que la distribution normale des caractéristiques humaines et la nécessité d'amplifier les différences entre les individus sont tout à fait inutiles dans le contexte scolaire. Pour évaluer les apprentissages, il faut examiner chaque apprenant par rapport à un résultat visé, indépendamment des autres de son groupe-classe. Mettre en évidence les changements ou les progrès individuels à un moment précis du cheminement scolaire génère des informations utiles à la régulation des apprentissages, souligne les forces, les faiblesses et les éléments à retravailler ou reconnaître les acquis. Même si l'interprétation critérielle des résultats, plutôt que leur interprétation normative, constitue le fondement de l'évaluation des apprentissages en salle de classe, il demeure tout aussi important que les résultats d'évaluation soient de qualité; il faut établir un moyen de s'assurer qu'ils soient fiables, vrais et utiles, pour qu'ils mènent à des jugements éclairés et à des décisions appropriées, tout cela en respectant la nature des apprentissages et l'utilisation faite de ces résultats.

À en juger par l'abondance d'écrits sur le sujet, cette problématique a alimenté la réflexion chez de nombreux experts et chercheurs. Ils ont proposé des manières de concevoir la démarche d'évaluation, fragmentant chaque étape en indiquant les aspects critiques de chacune d'elles, vulgarisé les différentes manières (normative, critérielle ou ipsative, avec ou sans seuil de réussite) d'interpréter les résultats et discuté des diverses fonctions (sommatrice, diagnostique ou formative) de l'évaluation. Tous les arguments au sujet de cette problématique ont fait ressortir la complexité de l'acte d'évaluer, tout en le rendant plus concret. Par la même occasion, ces échanges ont mis en relief les préoccupations fondamentales concernant la qualité des résultats d'évaluation en salle de classe. L'analyse de ces arguments a donné naissance au concept de justesse des résultats.

LA JUSTESSE DES RÉSULTATS

Le concept de justesse des résultats se base essentiellement sur la nécessité d'arriver à des résultats d'évaluation justes, tant en situation formative que sommatrice, à des jugements fondés et à des décisions éclairées. Sans prétendre pouvoir éclipser tout autre critère de qualité, le concept de justesse des résultats englobe toutes les étapes de la démarche d'évaluation et la situe par rapport au contexte d'apprentissage. Il souligne la nécessité de décrire les capacités de l'apprenant en termes de forces et de faiblesses, tout en se préoccupant de cohérence entre les expériences d'apprentissage vécues en salle de classe et les situations d'évaluation proposées.

En plus du sens d'exactitude, la notion de justesse des résultats englobe l'aspect d'équité. Cette dernière s'intéresse aux « cas d'espèce pour lesquels il n'est pas possible de poser un énoncé général qui s'y applique avec rectitude » (Aristote, trad. 1972, p. 267). La notion de justice fait abstraction de toute différence individuelle et réclame le traitement impartial de tous, alors que la justesse fait appel au jugement pour s'ajuster aux particularités des situations. Dans le domaine de l'évaluation, la notion de justice pourrait être associée aux concepts de validité et de fidélité, qui supposent l'application uniforme de mesures standardisées. La justesse des résultats, quant à elle, découlerait plutôt du respect raisonné des particularités, des conditions, du contexte et des besoins personnels des évalués, afin d'arriver à des résultats situant correctement chacun dans son développement.

Il existe plusieurs points critiques pour produire des résultats d'évaluation justes, mais aux fins de discussion, ils sont regroupés ici en quatre composantes.

I. Pertinence de la tâche d'évaluation par rapport aux habiletés visées

L'évaluation, tout comme l'enseignement, nécessite d'établir les habiletés visées par les activités pédagogiques. Que ces habiletés soient simples, comme le rappel de connaissances ou de faits, ou plus complexes comme le raisonnement, la production d'une œuvre ou la réalisation d'une tâche à partir de plusieurs habiletés interreliées, il est essentiel de les reconnaître et d'en établir les signes distinctifs. Ceci permet de ne pas les confondre avec d'autres habiletés semblables et d'en définir les dimensions importantes, soit les critères d'évaluation (Elliot, 1991; Glaser et Silver, 1994; Taylor, 1994). Cet aspect de la première composante du concept de justesse des résultats s'apparente à l'étape « intention » de la démarche d'évaluation, laquelle devrait donner tout son sens à la procédure d'évaluation engagée (Lussier, 1992).

La tâche d'évaluation permet l'émergence des comportements révélant l'état des apprentissages ciblés (Lane, 1993; Linn, Baker et Dunbar, 1991; Moss 1995; Popham, 1994). Sa longueur et son degré de difficulté s'ajustent donc aux caractéristiques des élèves. Qu'il s'agisse de choisir des réponses, d'en composer, d'effectuer une performance ou d'échanger verbalement, la tâche choisie valorise les habiletés ciblées et offre l'occasion de les apprécier, tout en faisant ressortir leur interaction s'il y a lieu (Mislevy, 1992). La stratégie adoptée, familière à l'élève, mais la tâche suffisamment nouvelle, informe de manière sûre au sujet des habiletés ciblées. De plus, des résultats d'évaluation justes proviennent d'une tâche signifiante pour l'évalué et compatible avec ses intérêts, définie clairement et dans un langage compréhensible pour lui (Gipps, 1994), ce qui justifie, à ses yeux, l'investissement de temps et d'énergie. Enfin, il va sans dire que la formulation de cette tâche ne comporte habituellement pas d'erreur technique.

Seules les activités d'apprentissage vécues précédemment par les élèves permettent de déterminer avec certitude les habiletés auxquelles la tâche

d'évaluation fait réellement référence. Comme l'expliquent Linn et al. (1991), une tâche peut aussi bien mettre en valeur des habiletés de mémorisation que des habiletés complexes, dépendant du degré de nouveauté de la tâche et des expériences vécues antérieurement. Il est primordial d'établir clairement les habiletés impliquées dans la tâche assignée et les indices qui révéleront leur manifestation. Une épreuve serait systématiquement non valide si elle pouvait être réussie avec des stratégies d'apprentissage élémentaires, comme des automatismes, sans comprendre les principes (Lane, 1993), par exemple, résoudre un problème presque en tout point identique à un autre résolu en classe quelques jours avant.

Des critères bien choisis et bien définis aident à juger une performance et communiquent explicitement ce qui est attendu. Ils sont propres à l'activité en milieu réel et applicables à une variété de tâches semblables. De tels critères peuvent s'ajuster au niveau de développement des évalués et à leur cheminement scolaire; leur nombre varie selon la complexité de la performance. Dans une situation d'évaluation formative, les critères choisis pour juger la performance sont spécifiques à divers aspects des apprentissages visés; ils servent à repérer les erreurs de parcours et à mieux saisir la manière de penser de l'élève, puisque le but poursuivi est d'encourager ses efforts, de mieux cibler les activités d'apprentissage subséquentes ou de consolider les acquis. Dans le cas d'une situation d'évaluation sommative, des critères plus englobants permettent de poser un regard critique sur les acquis, en comparaison avec les résultats visés et les normes établies.

L'agencement optimal des habiletés visées et le type de tâche choisi pour les faire ressortir s'avère crucial (Stiggins, 1994; Taylor, 1994); c'est le fondement de l'obtention de résultats d'évaluation justes. Des lacunes à ce niveau-ci de la démarche d'évaluation mettraient en péril la justesse des résultats, peu importe la rigueur appliquée aux autres composantes du concept.

II. Cohérence de l'évaluation avec l'activité pédagogique

Pour être intégrée à l'apprentissage, l'activité d'évaluation doit s'insérer convenablement dans l'activité pédagogique (Lane, 1993) et arriver au moment opportun, c'est-à-dire une fois que l'apprenant a eu suffisamment d'occasions de développer les apprentissages en question (Zahorik, 1994). Même si toute situation d'évaluation présente habituellement un aspect nouveau, il est essentiel que les apprenants soient familiers avec le genre de situation d'évaluation proposée (Gardner, 1992), bénéficient d'une mise en situation convenable et puissent demander des précisions s'ils ne comprennent pas bien la tâche à réaliser. Non seulement doivent-ils connaître les critères qui guident l'évaluation, mais ils doivent aussi avoir déjà eu l'occasion de les appliquer (Lussier et Turner, 1995; Shepard, 1989). De plus, l'évaluation étant habituellement menée dans le but de servir l'apprentissage ou de reconnaître les acquis, les résultats communiquent à l'apprenant une appréciation de sa performance. Les résultats réfèrent donc à

des forces et à des faiblesses, avec un degré de spécificité ajusté au type de réinvestissement envisagé : plus détaillés et plus spécifiques pour un réinvestissement immédiat, plus globaux et plus généraux pour un réinvestissement à plus long terme. Pour l'élève insatisfait d'un premier résultat, la réévaluation offre une occasion de se reprendre et de progresser.

III. Transparence de l'activité évaluative

La transparence d'une démarche d'évaluation se manifeste par la participation, la discussion et l'appropriation par l'élève des divers aspects de l'évaluation de ses apprentissages, dans le but de s'assurer qu'il sait ce qui est attendu de lui et l'utilisation qui sera faite des résultats. Selon son niveau de maturité, l'élève peut aider à établir les critères d'évaluation et expérimenter leur application à ses activités d'apprentissage; l'autoévaluation et l'évaluation par les pairs à des fins formatives s'avèrent particulièrement propices à cette prise en charge (Doyon et Juneau, 1991). L'apprentissage graduel d'une plus grande autonomie responsabilise et agit indirectement sur la motivation. La présentation commentée de travaux de diverses qualités aide à communiquer clairement les attentes et les critères d'évaluation et favorise le transfert de ces critères à d'autres activités d'apprentissage (Taylor, 1994).

Les mesures de transparence développent une meilleure communication entre les enseignants et les élèves (Linn et al., 1991), contribuent à réduire la méfiance et l'anxiété face à une situation d'évaluation (Wiggins, 1993), à assurer la justesse des résultats et à favoriser un rendement optimal. La description des critères, éléments cruciaux de cette transparence, appuie aussi le jugement porté sur les réalisations des élèves, favorisant leur traitement objectif et constant de la part de l'enseignant (Glaser et Silver, 1994). Le résultat, formulé en termes de progrès, de difficultés, de maîtrise ou de réussite, selon la visée de l'évaluation, se rattache à la performance fournie et aux critères qui ont servi à la juger. Le retour en classe sur l'activité d'évaluation donne l'occasion à l'élève d'exprimer son accord ou son point de vue sur la performance fournie et le résultat obtenu.

IV. Absence de désavantage circonstanciel

Malgré les nombreuses mises en garde décrites jusqu'à maintenant, il existe encore d'autres obstacles à la justesse des résultats. Même s'il n'est pas possible de tout contrôler, il faut au moins être conscient de ce qui fausse les résultats obtenus. Par exemple, à cause de l'absence de limite de temps imposé, les devoirs faits à la maison peuvent d'une certaine manière témoigner avec plus de justesse des capacités d'un élève; par contre, puisqu'il n'est pas possible d'affirmer hors de tout doute que l'élève a fait le travail lui-même, cette stratégie réduit la justesse des résultats (Leder, 1992). La situation d'évaluation doit aussi rassembler des conditions favorables à l'élève (Pellegrino, 1992); le bruit excessif,

les limites de temps, la fatigue, des événements excitants, nuisent à une réalisation optimale et contribuent à l'instabilité de la performance observée (Dunbar, Koretz et Hoover, 1991). L'enseignant doit aussi demeurer conscient des effets possiblement néfastes de la fatigue, de ses préférences personnelles et de la connaissance du rendement antérieur de chacun sur l'objectivité de son jugement.

Une stratégie choisie peut aussi être plus favorable à certains élèves qu'à d'autres, dépendant de leurs caractéristiques personnelles (Lussier et Turner, 1995). Par exemple, comparés à des élèves sociables et sûrs d'eux-mêmes, les élèves anxieux et timides sont désavantagés dans leur performance lors d'évaluations orales. De manière semblable, un élève écrivant avec difficulté est plus à l'aise face à un questionnaire à réponses choisies que face à un examen exigeant des réponses élaborées. De plus, une tâche qui ne comporte pas de préjugé culturel ou sexiste produirait des résultats d'évaluation plus justes (Peddie, 1992). Cette qualité optimale d'une situation d'évaluation étant difficile à atteindre, tout résultat d'évaluation insatisfaisant devrait mener à conclure non pas au manque de maîtrise des savoirs évalués, mais à l'incapacité de démontrer la maîtrise; l'apprentissage n'est pas nécessairement en cause (Traub, 1990).

* * * * *

Les quatre composantes de la justesse des résultats d'évaluation sont donc la pertinence de la tâche par rapport aux habiletés visées, la cohérence avec l'activité pédagogique, la transparence de l'activité évaluative et l'absence de désavantage circonstanciel. Ainsi, la première composante du critère de justesse est l'établissement des habiletés à évaluer et le choix d'une stratégie susceptible de bien les mettre en valeur. Les trois autres composantes contribuent à mettre en lumière les apprentissages réalisés. La cohérence avec l'activité pédagogique (deuxième composante) et la transparence de l'activité évaluative (troisième composante) sont du même ordre, se déroulant souvent simultanément. En dernier lieu, l'absence de désavantage circonstanciel regroupe un ensemble de circonstances à éviter lors de l'apprentissage ou de son évaluation, afin de ne pas réduire la qualité des résultats.

Les critères d'évaluation sont reliés de près à l'établissement des habiletés ciblées, mais leur communication aux élèves relève de la transparence et leur application, de l'activité pédagogique. Il en est de même de la communication des résultats en termes de forces et de faiblesses. Elle doit être pressentie lors du choix de la tâche, permettre un réinvestissement dans l'activité pédagogique et donner de la transparence, en communiquant à l'élève en termes concrets une appréciation de ses réalisations.

CONCLUSION

Le critère de justesse des résultats répond à deux préoccupations actuelles en évaluation des apprentissages : d'abord, il constitue un modèle cohérent avec la

nature même des apprentissages; de plus, il convient à différents types de stratégies d'évaluation, à différents buts, différentes disciplines et différents niveaux scolaires. En mettant en valeur des principes fondamentaux d'évaluation des apprentissages, en faisant ressortir les points stratégiques sur lesquels il faut se concentrer et les failles à éviter, il pourrait favoriser des pratiques d'évaluation appropriées et intégrées au processus pédagogique; il permettrait d'uniformiser les exigences et les contextes d'évaluation selon les normes à atteindre. En donnant plus de rigueur aux démarches d'évaluation entreprises en salle de classe, le concept de justesse contribuerait à produire des résultats d'évaluation justes, c'est-à-dire des résultats d'évaluation qui situent correctement l'apprenant dans son apprentissage, déterminent ses forces et ses faiblesses, ce qu'il maîtrise ou ne maîtrise pas. Ainsi un résultat juste sera-t-il fondé, sensé, représentatif, signifiant et vrai, entaché d'un minimum d'erreur.

Cette conception de la qualité d'un résultat d'évaluation des apprentissages en salle de classe paraît appropriée mais tant qu'elle n'est pas confrontée à la réalité par une mise à l'essai rigoureuse en milieu scolaire, elle demeure au stade d'hypothèse de travail. Le présent cadre conceptuel de la qualité des résultats d'évaluation des apprentissages en salle de classe a donc fait l'objet d'une recherche empirique. Le déroulement de cette mise à l'essai et ses résultats seront décrits dans une publication ultérieure.

RÉFÉRENCES

- Aristote. (1972). *Éthique à Nicomaque* (3^e éd., nouvelle traduction avec introduction, notes et index par J. Tricot). Paris : Vrin.
- Bachor, D. et Anderson, J. (1994). Elementary teachers' assessment practices as observed in the province of British Columbia. *Assessment in Education*, 1, 63–93.
- Baron, J. (1991). Performance assessment: Blurring the edges of assessment, curriculum, and instruction. Dans G. Kulm et S. Malcom (dirs), *Science assessment in the service of reform* (p. 247–266). Washington, DC : American Association for the Advancement of Science.
- Bateson, D. (1990). Measurement and evaluation practices of British Columbia science teachers. *Alberta Journal of Educational Research*, 36, 45–51.
- Bercier-Larivière, M. (1994). *La motivation des élèves de 6^e, 7^e et 8^e année et leur perception des pratiques d'évaluation des apprentissages*. Thèse de maîtrise non publiée, Université d'Ottawa, Ottawa.
- Bercier-Larivière, M. (1998). *La justesse des résultats, critère de qualité de l'évaluation des apprentissages en salle de classe*. Thèse de doctorat non publiée, Université d'Ottawa, Ottawa.
- Biggs, J. (1995). Assessing for learning: Some dimensions underlying new approaches to educational assessment. *Alberta Journal of Educational Research*, 41, 1–17.
- Cardinet, J. (1985). Des instruments d'évaluation pour chaque fonction. *Mesure et évaluation en éducation* 8(1–2), 45–118.
- Conseil supérieur de l'éducation. (1992). *Évaluer les apprentissages au primaire : Un équilibre à trouver*. Québec : Conseil supérieur de l'éducation.
- Cox, K. I. (1994). What "counts" in English class? A study of classroom assessment practices among

- California high school teachers (Thèse de doctorat, University of California, Riverside, 1994). *Dissertation Abstracts International*, 56(01), 851A.
- Davinroy, K. H. (1994, avril). "How does my teacher know what I know?" Third graders' perceptions of math, reading, and assessment. Communication présentée à la réunion annuelle de l'American Educational Research Association (AERA), New Orleans.
- DeLandsheere, G. (1992). *Dictionnaire de l'évaluation et de la recherche en éducation*. Paris : Presses universitaires de France.
- Dempster, F. (1992). Using tests to promote learning: A neglected classroom resource. *Journal of Research and Development in Education*, 25, 213–217.
- Doyon, C. et Juneau, R. (1991). *Faire participer l'élève à l'évaluation de ses apprentissages*. Laval : Éditions Beauchemin.
- Dunbar, S., Koretz, D. et Hoover, H. (1991). Quality control in the development and use of performance assessments. *Applied Measurement in Education*, 4, 289–303.
- Elliot, S. (1991). Authentic assessment: An introduction to a neobehavioral approach to classroom assessment. *School Psychology Quarterly*, 6, 273–278.
- Forgette-Giroux, R., Simon, M. et Bercier-Larivière, M. (1996). Les pratiques d'évaluation des apprentissages en salle de classe : perceptions des enseignants et des enseignantes. *Revue canadienne de l'éducation*, 21, 384–395.
- Fraser, B. et Tobin, K. (1991). Combining qualitative and quantitative methods in classroom environment research. Dans B. Fraser et H. Walberg (dirs), *Educational environments: Evaluation, antecedents and consequences* (p. 271–292). New York : Pergamon Press.
- Gardner, H. (1992). Assessment in context: The alternative to standardized testing. Dans B. Gifford et M. O'Connor (dirs), *Changing assessments: Alternative views of aptitude, achievement, and instruction* (p. 77–120). Boston : Kluwer.
- Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. Washington, DC : Falmer Press.
- Gipps, C. et Murphy, P. (1994). *A fair test? Assessment, achievement and equity*. Philadelphia : Open University Press.
- Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*, 18, 519–521.
- Glaser, R., et Silver, E. (1994). *Assessment, testing and instruction: Retrospect and prospect* (Center for the Study of Evaluation Tech. Rep. No. 379). Los Angeles : National Center for Research on Evaluation, Standards, and Student Testing.
- Harnisch, D. et Mabry, L. (1993). Issues in the development and evaluation of alternative assessments. *Journal of Curriculum Studies*, 25, 179–187.
- Herman, J. et Dorr-Bremme, D. (dirs). (1984). *Teachers and testing: Implications from a national study*. Los Angeles : University of California at Los Angeles, Center for the Study of Evaluation.
- Herman, J. L. et Klein, D. (1996). Evaluating equity in alternative assessment: An illustration of opportunity-to-learn issues. *Journal of Educational Research*, 89, 246–256.
- Higgins, N. et Rice, E. (1991). Teachers' perspectives on competency-based testing. *Educational Technology Research and Development*, 39(3), 59–69.
- Hipps, J. (1993, avril). *Trustworthiness and authenticity: Alternate ways to judge authentic assessments*. Communication présentée à la réunion annuelle de l'American Educational Research Association (AERA), Atlanta.

- Kulm, G. (1994). *Mathematics assessment: What works in the classroom*. San Francisco : Jossey-Bass.
- Kusch, J. W. (1995). Teaching in a surveillance mode: A case study of how methods course syllabi and practicum classrooms construct assessment as classroom control (Thèse de doctorat, University of Wisconsin-Madison, 1995). *Dissertation Abstracts International*, 56(05), 1645A.
- Lane, S. (1993). The conceptual framework for the development of a mathematics performance assessment instrument. *Educational Measurement: Issues and Practice*, 12(2), 16–23.
- Leder, G. (1992). *Assessment and learning of mathematics*. Victoria, Australia : Australian Council for Educational Research.
- Linn, R., Baker, E. et Dunbar, S. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20(8), 15–21.
- Lussier, D. (1992). *Évaluer les apprentissages, dans une approche communicative*. Paris : Hachette.
- Lussier, D. et Turner, C. (1995). *Le point sur... L'évaluation en didactique des langues*. Anjou : Centre Éducatif et Culturel.
- Macdonald, W. J. (1993). Student assessment in the elementary intermediate visual arts program: A closer look at teacher assessment methods (Thèse de doctorat, University of Toronto, 1993). *Dissertation Abstracts International*, 54(09), 3299A.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5–11.
- Ministère de l'Éducation et de la Formation de l'Ontario. (1997). *Le curriculum de l'Ontario de la 1^{re} à la 8^e année. Français*. Toronto : Auteur.
- Mislevy, R. (1992). *Linking educational assessments: Concepts, issues, methods, and prospects*. Princeton : Educational Testing Service.
- Moss, P. (1995). Themes and variations in validity theory. *Educational Measurement: Issues and Practice*, 14(2), 5–13.
- Peddie, R. (1992). *Beyond the norm? An introduction to standards-based assessment*. Wellington : New Zealand Qualifications Authority.
- Pellegrino, J. (1992). Commentary: Understanding what we measure and measuring what we understand. Dans B. Gifford et M. O'Connor (dirs), *Changing assessments: Alternative views of aptitude, achievement, and instruction* (p. 275–300). Boston : Kluwer.
- Popham, J. (1990). *Modern educational measurement: A practitioner's perspective* (2^e éd.). Toronto : Prentice-Hall.
- Popham, J. (1994). The instructional consequences of criterion-referenced clarity. *Educational Measurement: Issues and Practice*, 13(4), 15–18, 30.
- Pryor, E. L. (1992). Case studies of three first-grade teachers' assessment and evaluation of literacy (grading) (Thèse de doctorat, Kent State University, 1992). *Dissertation Abstracts International*, 56(08), 2669A.
- Rogers, T. (1991). Educational assessment in Canada: Evolution or extinction? *Alberta Journal of Educational Research*, 37, 179–192.
- Shepard, L. (1989). Why we need better assessments. *Educational Leadership*, 46(7), 4–9.
- Shepard, L. (1992). Commentary: What policy makers who mandate tests should know about the new psychology on intellectual ability and learning. Dans B. Gifford et M. O'Connor (dirs), *Changing assessments: Alternative views of aptitude, achievement, and instruction* (p. 301–328). Boston : Kluwer.

- Stiggins, R. (1990). Toward a relevant classroom assessment research agenda. *Alberta Journal of Educational Research*, 36, 92–97.
- Stiggins, R. (1991). Assessment literacy. *Phi Delta Kappan*, 72, 534–539.
- Stiggins, R. (1994). *Student-centered classroom assessment*. Toronto : Maxwell Macmillan.
- Stiggins, R. J. et Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. New York : State University of New York Press.
- Taylor, C. (1994). Assessment for measurement or standards: The peril and promise of large-scale assessment reform. *American Educational Research Journal*, 31, 231–262.
- Tobin, K., Butler-Kahle, J. et Fraser, B. (dirs). (1990). *Windows into science classrooms*. New York : Palmer Press.
- Trahan, M. et Dassa, C. (1978). La courbe normale en éducation : postulat, constante naturelle ou erreur scientifique. *Revue des sciences de l'éducation*, 4, 71–80.
- Traub, R. (1990). Assessment in the classroom: What is the role of research? *Alberta Journal of Educational Research*, 36, 85–91.
- Wiggins, G. (1993). *Assessing student performance*. San Francisco : Jossey-Bass Publishers.
- Wilson, R. (1989). Evaluating student achievement in an Ontario high school. *Alberta Journal of Educational Research*, 35, 134–144.
- Wilson, R. (1990). Classroom processes in evaluating student achievement. *Alberta Journal of Educational Research*, 36, 4–17.
- Zahorik, J. (1994). Teacher evaluating behavior. Dans T. Husen et T. Postlewaite (dirs), *International Encyclopedia of Education* (2^e éd., Vol. 10, p. 6012–6015). Toronto : Pergamon.

Micheline Bercier-Larivière est Chef, Imputabilité, Conseil des écoles catholiques de langue française du Centre-Est de l'Ontario, 4000, rue Labelle, Gloucester (Ontario) K1J 1A1. Renée Forgette-Giroux est professeure à la Faculté d'éducation, Université d'Ottawa, 145 Jean-Jacques Lussier, Ottawa (Ontario) K1N 6N5.